Future Timelines: Extraction and Visualization of Future-related Content From News Articles

Juwal Regev
juwal.regev@student.uibk.ac.at
University of Innsbruck
Innsbruck, Austria

Adam Jatowt
University of Innsbruck
Innsbruck, Austria
adam.jatowt@uibk.ac.at

Michael Färber
Karlsruhe Institute of Technology
Karlsruhe, Germany
michael.faerber@kit.edu

ABSTRACT

In today's rapidly evolving world, maintaining a comprehensive overview of the future landscape is essential for staying competitive and making informed decisions. However, given the overwhelming volume of daily news, manually obtaining a thorough overview of an entity's future prospects is quite challenging. To address this, we present a system that is designed to automatically extract future-related information of a queried entity from news articles. Our approach involves fine-tuning a language model on a novel multi-source dataset consisting of 6,800 annotated sentences to identify future-related sentences. We use topic modeling to extract the main topics from the data, which, along with their contents, are subsequently ranked by relevance. A temporal tagger is utilized to detect temporal expressions and map them to specific dates. This enables the sentences to be presented on an interactive timeline. User evaluations have shown favorable feedback on the timelines and summaries our system produces.

CCS CONCEPTS

• Information systems → Language models; **Document topic models**; **• Computing methodologies → Information extraction**.

KEYWORDS

Future-related Content Extraction, Sentence Classification, Topic Modeling, Time-Tagging, Timeline Generation, Natural Language Processing

ACM Reference Format:
Juwal Regev, Adam Jatowt, and Michael Färber. 2023. Future Timelines: Extraction and Visualization of Future-related Content From News Articles. In *Proceedings of The 17th ACM International Conference on Web Search and Data Mining (WSDM '24).* ACM, New York, NY, USA, 4 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

A vast amount of news articles is published on the web every single day. A significant amount of those articles contains information that is predictive or related to future events. Extracting and analyzing such information regarding a specific entity would offer a comprehensive overview of its future prospects.

However doing this manually is not feasible as one would have to sift through the articles line by line, identifying nuanced references to the future. Furthermore, this would have to be done regularly so as to always have the most up-to-date information. Thus, it is clear that an automated system, capable of extracting, processing and visualizing this information is needed. This system would be beneficial in a variety of different fields. From analyzing predictions related to stock market movements, monitoring corporate developments as well as gaining insights into market trends and emerging technologies, the potential applications are limitless.

To address this problem, we present a system specifically designed for extracting, summarizing, ranking and visualizing future-related content within news collections. The process is initiated by a user providing an entity, whose future should be explored. The system then downloads and preprocesses thousands of news articles related to the provided entity. In the next step, a neural network classifier is employed to classify sentences as either being future-related or not. Following classification, these sentences are organized into distinct topics, which are then presented to the user in a segmented manner. Additionally, the topics are labeled with representative keywords that provide insight into the content of each cluster. To provide an even better overview, we incorporate a temporal perspective into the presentation of the results. We do this by analyzing the sentences for temporal expressions, extracting and normalizing them to a date and then mapping the sentences, ranked by their relevance, onto a timeline.

In the following, we will delve into the system's architecture, providing implementation details, explaining the user interface and presenting the findings of a user study conducted to assess the system's result quality.

2 RELATED WORK

Automatic identification and extraction of future-related information from text has been researched for decades. Baeza-Yates [1] first formalized the concept of "future retrieval". Early methods relied on time-taggers and predefined temporal expressions to extract data, but this resulted in limited diversity [3–5, 7, 8]. Data was analyzed for unique properties [6] to create features for classifier training [11]. For example, one study uses morphosemantic patterns as features [10], while another classifies clauses using features like POS tags and word co-occurrences to determine if a sentence refers to

WSDM '24, March 04–08,2024, Mérida, MX

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

the future [16]. Some systems also predict future events using past data [12, 13]. In Timeline Summarization, content is clustered and ranked to form a timeline. Techniques involve using TF-IDF-vectors or SBERT embeddings [14] to represent sentences, and methods like Affinity Propagation [15, 17] and Markov Clustering [2] for clustering. Sentence rankings inside of a topic or event often depend on the similarity to a centroid sentence [2] or on metrics like date-importance and informativeness [15].

3 SYSTEM

In the following sections, we will introduce our publicly available[1] dataset and system and explore each of its components in depth, gaining a detailed understanding of the overall workflow and examining implementation details.

3.1 Dataset

As already mentioned, our approach to data gathering differs from traditional methods. Earlier strategies have frequently relied on extracting data using temporal expressions. However, this approach has its limitations. Predictions can often be intricate, lacking explicit dates or simple temporal cues. To address this, our dataset incorporates a rich variety of 6,800 manually labeled sentences. They have been collected from different sources without relying on specific queries, and therefore they exhibit unique lexical and structural features and contain a diverse set of topics. This enables our classifier to identify future-related content that does not contain standard temporal expressions hinting at future references. Table 1 provides a breakdown of our data sources.

Sources	Positives	Negatives	Sentences without temporal expressions (%)
Longbets	448	0	20
Horizons	51	62	84
ChatGPT	305	305	70
News	2501	3128	39
Total	**3305**	**3495**	**41**

Table 1: Data Source Analysis: Count of Future-Related (Positive) vs. Non-Future-Related (Negative) Sentences, along with Percentage Free from Temporal Expressions.

Longbets[2] is a platform that enables users to share and bet on predictions. In contrast, *Horizons*[3] is more focused on emerging technologies. The *News* sentences were extracted from New York Times[4] articles. *ChatGPT*[5] was used for its capability to generate highly specific data. Using prompts like *"Generate 50 sentences about various topics that contain future related information. You should use words like 'could, 'may', 'might', ..."*, we were able to obtain highly specific and tailored training data.

3.2 System Overview

We will now take a closer look at our system's five main components: 1) Data Retrieval and Preprocessing, 2) Classification, 3) Topic Modeling, 4) Postprocessing, 5) Time-Tagging. A visualization of the workflow can be seen in Figure 2.

3.2.1 Fetching and Preprocessing. To initiate the process, the user has to provide a query to the system, which can take the form of a topic, an entity, or any other input whose future should be analyzed. The user has the option to choose the number of articles that should be downloaded, as well as the time frame from which these articles should be fetched. Once the articles are downloaded, they are split up into their constituent sentences. These sentences subsequently undergo preprocessing to prepare them for the next steps.

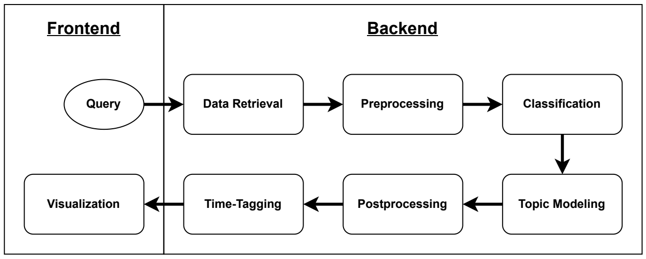


Figure 2: Schematic Overview of the System's Data Processing Pipeline.

3.2.2 Classification. Our classifier was implemented using a DistilRoBERTa[6] model, which was fine-tuned by appending a feed-forward network to the output. DistilRoBERTa is a smaller, faster version of the renowned RoBERTa language model [9], trained to preserve its performance on downstream tasks. It was chosen over the larger RoBERTa model because our system requires real-time inference, where speed is critical. To further increase performance, the sequence length was reduced to 50 tokens. Considering that the model only processes one sentence at a time, and given that the average sentence length in our dataset is 20 words, this is sufficient. The outputs from the DistilRoBERTa model undergo mean pooling to produce a 768-dimensional vector, which is then fed into our fine-tuning feed-forward network.

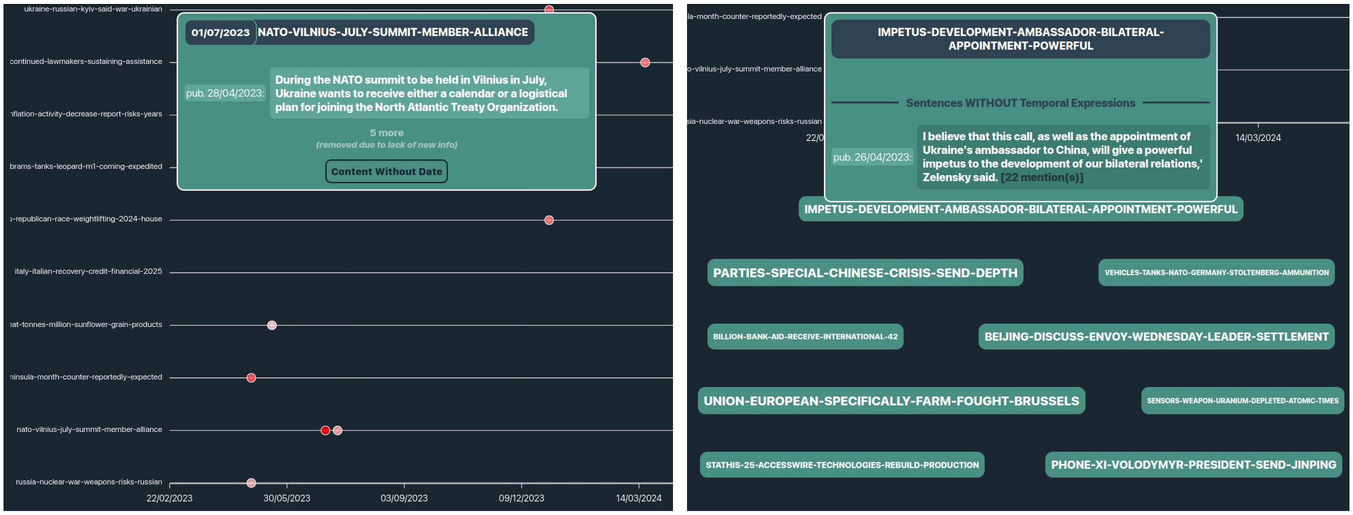
Training Insights. We fine-tuned the network on our dataset over the course of 14 epochs with the following configurations:

- Batch size: 8
- Learning rate: 1.5e-6
- Warmup: 0.2
- Weight decay: 0.001

Validation Insights. We employed 10-fold cross-validation to obtain performance estimates for the model. The following metrics show the average scores we achieved:

- Accuracy: 0.965
- Precision: 0.953
- Recall: 0.98
- F1-score: 0.97
- AUC-ROC: 0.964

[1] https://github.com/regevson/chronicle2050/
[2] https://longbets.org/
[3] https://radar.envisioning.io/horizons/
[4] https://www.nytimes.com/
[5] https://chat.openai.com/
[6] https://huggingface.co/distilroberta-base


ukraine-russian-kyiv-said-war-ukrainian
continued-lawmakers-sustaining-assistance
nflation-activity-decrease-report-risks-years
brams-tanks-leopard-m1-coming-expedited
;-republican-race-weightlifting-2024-house
italy-italian-recovery-credit-financial-2025
eat-tonnes-million-sunflower-grain-products
ninsula-month-counter-reportedly-expected
nato-vilnius-july-summit-member-alliance
russia-nuclear-war-weapons-risks-russian
22/02/2023
30/05/2023
03/09/2023
09/12/2023
14/03/2024
01/07/2023
NATO-VILNIUS-JULY-SUMMIT-MEMBER-ALLIANCE
pub. 28/04/2023:
5 more
(removed due to lack of new info)
Content Without Date
la-month-counter-reportedly-expected
o-vilnius-july-summit-member-alliance
sia-nuclear-war-weapons-risks-russian
22/0
14/03/2024
IMPETUS-DEVELOPMENT-AMBASSADOR-BILATERAL-APPOINTMENT-POWERFUL
Sentences WITHOUT Temporal Expressions
pub. 26/04/2023:
[22 mention(s)]
IMPETUS-DEVELOPMENT-AMBASSADOR-BILATERAL-APPOINTMENT-POWERFUL
PARTIES-SPECIAL-CHINESE-CRISIS-SEND-DEPTH
VEHICLES-TANKS-NATO-GERMANY-STOLTENBERG-AMMUNITION
BILLION-BANK-AID-RECEIVE-INTERNATIONAL-42
BEIJING-DISCUSS-ENVOY-WEDNESDAY-LEADER-SETTLEMENT
UNION-EUROPEAN-SPECIFICALLY-FARM-FOUGHT-BRUSSELS
SENSORS-WEAPON-URANIUM-DEPLETED-ATOMIC-TIMES
STATHIS-25-ACCESSWIRE-TECHNOLOGIES-REBUILD-PRODUCTION
PHONE-XI-VOLODYMYR-PRESIDENT-SEND-JINPING

During the NATO summit to be held in Vilnius in July, Ukraine wants to receive either a calendar or a logistical plan for joining the North Atlantic Treaty Organization.

I believe that this call, as well as the appointment of Ukraine's ambassador to China, will give a powerful impetus to the development of our bilateral relations,' Zelensky said.

Figure 1: Visualization of the system's key components: (left) The interactive timeline featuring an open tooltip section, and (right) the Word Cloud also with an open tooltip section.

Inference. The preprocessed sentences are fed into the fully trained model, which outputs a score between 0 and 1, indicating the probability that the sentence is future-related. Sentences with a probability above 0.9 are considered positives and are forwarded to the topic modeling step.

3.2.3 Topic Modeling.

To present the user with sentences assigned to topics, we perform topic modeling using BERTopic[7], a system that extracts topics from text documents by embedding them in a high-dimensional space using SBERT[8] and then clustering the embeddings by similarity. It also extracts representative keywords from the topics, which are used as topic headings. We configured BERTopic to perform dimensionality reduction with UMAP[9] and clustering with HDBSCAN[10].

3.2.4 Postprocessing.

To improve the quality of topic modeling results and rank sentences based on their relevance, we postprocess the results in two ways. First, we remove outliers from topics by matching the topic keywords against the words in each sentence. If a sentence contains fewer than three topic keywords, it is removed from the dataset. The next step eliminates redundancy within a topic and evaluates the relevance of each sentence. Using the previously generated SBERT sentence embeddings from the topic modeling step, we calculate the cosine similarity between all sentence pairs within the same topic. When two sentences have a cosine similarity above 0.8, they are considered duplicates. The shorter sentence is discarded, and the duplicate count of the longer sentence is incremented. This duplicate count serves as a measure of sentence relevance. The greater the number of duplicates, the more frequently the sentence appeared in news articles and the more relevant it might be.

3.2.5 Time-Tagging.

To create a timeline, we need to map sentences to concrete dates. This requires the identification and resolution of temporal expressions in the sentences. The SUTime[11] time-tagger was employed for this purpose. It detects temporal information and resolves it to the referenced date, incorporating the article's publication date. Some rules had to be added to ensure that all temporal expressions are mapped to dates, such as mapping *'Summer'* to *'07.08'*. While the tagger generally performs well, there are some problems with ambiguous references like *'On Tuesday, he revealed his decision, which will become official soon'*. The time-tagger maps *"on Tuesday"* to the coming Tuesday, which is in the future, even though the sentence is referring to a past event. This is a known limitation of SUTime, which is mitigated by simply discarding extracted temporal expressions that reference weekdays.

3.3 Demonstration System

The system was implemented as a web app with a Vue.js[12] frontend and a Django[13] backend. It can be accessed with a standard browser at the following URL: *http://chronicle2050.regevson.com*. Due to the workflow taking a significant amount of time, the whole process is executed multiple times, each time downloading new articles, classifying them and then clustering them together with the previously extracted sentences. When new results are ready, a button appears, which lets the user give permission to load in the updated data.

3.3.1 Settings.

Upon entry, the website shows a settings panel that provides the user with some adjustable configurations. Here, decisions about whether updates should be automatic or manual can be made. There's also the possibility to specify the quantity of articles and their time frame. A text field prompts users to specify an entity for exploration.

[7] https://maartengr.github.io/BERTopic/index.html
[8] https://www.sbert.net/
[9] https://umap-learn.readthedocs.io/en/latest/
[10] https://hdbscan.readthedocs.io/en/latest/
[11] https://nlp.stanford.edu/software/sutime.shtml
[12] https://vuejs.org/
[13] https://www.djangoproject.com/

3.3.2 *Timeline.*

The website's core component is the timeline visualization. As shown in Figure 1 (left), the y-axis shows the different topics, whereas the x-axis displays future dates referenced by the sentences. A data point on the timeline represents a collection of sentences belonging to the same topic, that all reference the same date. This collection of sentences is revealed when hovering over the datapoint in the form of a tooltip section. They are ranked according to their relevance score, which is also displayed next to each sentence. There is also an option to delve into the full article for more context, by simply clicking on a sentence. The data point itself has a color assigned to it, representing the relevance of the sentences it contains, with red colors indicating higher relevance.

3.3.3 *Word Cloud.*

Complementing the timeline, the word cloud visualization, depicted in Figure 1 (right), provides the user with sentences that do not contain any temporal expressions. Each topic is represented as a cloud, whose size corresponds to the mean relevance of the contained sentences. On interaction, these clouds unfold to reveal the associated sentences, again along with their relevance scores and with links to their origin articles.

4 EVALUATION

To assess the effectiveness of our system, we developed a six-task exercise and presented it to six different users. Each user had to complete the exercise with one of three entities. These tasks were designed to examine the timeline and word cloud components. The tasks included: 1) summarizing the entity's future outlook, 2) determining which topic and date have the highest/lowest relevance and analyzing the overall ranking, 4) engaging with the timeline to find incorrectly resolved temporal expressions, 5) summarizing the content of the most relevant word cloud, 6) examining the word clouds for duplicate content and 7) examining both components for sentences without future-related content.

The users provided lengthy summaries, implying a wide range of predictions for upcoming years. The color-coding of datapoints managed to successfully convey relevance, but users noticed that nearer predictions were generally marked as more relevant, while key distant predictions often had lower relevance scores. Some users found this to be a problem, as distant events are often more significant and require more time to unfold. The time-tagging was generally accurate but struggled with some past predictions like *"In 2019, he correctly predicted that a pandemic would occur by the end of the year"*. This is because the tagger resolves temporal expressions relative to the parent article's publishing date, and does not understand that *"in 2019"* should be the anchor date. Some users found similar sentences in the same cluster, but the information they provided was slightly different. This provided additional context, which was well-received.

Overall, the system received strong approval from its users. On a five-star scale, the topic and classification quality received an average rating of 4 stars, while time-tagging quality received 3 stars. The ranking was evaluated at 3.5 stars, and the visualization received a 4.5-star average. Users were particularly satisfied with the ability of the system to provide information about an entity's future plans, events, and decisions. However, the biggest critique was the system's speed, especially the time it takes to retrieve the initial results. This should be kept for future work.

5 CONCLUSIONS

By developing a multi-source dataset of 6,800 labeled sentences, we fine-tuned a DistilRoBERTa model to identify future references. We then used BERTopic for topic modeling and SUTime for time-tagging to extract topics from the data and detect and resolve temporal expressions. A dedicated postprocessing step improved cluster quality and ranked sentences according to their relevance. We also designed an intuitive interface that presents the results with an interactive timeline and word clouds. A user study confirmed the system's ability to offer a detailed overview of the future, while also highlighting potential areas for improvement, especially speed.

REFERENCES

[1] Ricardo A. Baeza-Yates. [n. d.]. Searching the Future.

[2] Demian Gholipour Ghalandari and Georgiana Ifrim. 2020. Examining the State-of-the-Art in News Timeline Summarization. In *Proc. of the 58th ACL*. Association for Computational Linguistics, Online, 1322–1334. https://doi.org/10.18653/v1/2020.acl-main.122

[3] Adam Jatowt and Ching-man Au Yeung. [n. d.]. Extracting Collective Expectations about the Future from Large Text Collections. In *Proc. of the 20th CIKM* (Glasgow Scotland, UK, 2011-10-24). ACM, 1259–1264. https://doi.org/10.1145/2063576.2063759

[4] Adam Jatowt, Kensuke Kanazawa, Satoshi Oyama, and Katsumi Tanaka. 2009. Supporting Analysis of Future-Related Information in News Archives and the Web. In *Proc. of the 9th ACM/IEEE-CS Joint Conference on Digital Libraries* (Austin, TX, USA) *(JCDL '09)*. Association for Computing Machinery, New York, NY, USA, 115–124. https://doi.org/10.1145/1555400.1555420

[5] Adam Jatowt, Hideki Kawai, Kensuke Kanazawa, Katsumi Tanaka, Kazuo Kunieda, and Keiji Yamada. 2010. Analyzing collective view of future, time-referenced events on the web. 1123–1124. https://doi.org/10.1145/1772690.1772835

[6] Kensuke Kanazawa, Adam Jatowt, and Katsumi Tanaka. [n. d.]. Improving Retrieval of Future-Related Information in Text Collections. In *IEEE/WIC/ACM 2011* (2011), Vol. 1. 278–283. https://doi.org/10.1109/WI-IAT.2011.250

[7] Nattiya Kanhabua, Roi Blanco, and Michael Matthews. [n. d.]. Ranking Related News Predictions. In *Proc. of the 34th ACM SIGIR* (New York, NY, USA, 2011) *(SIGIR '11)*. Association for Computing Machinery, 755–764. https://doi.org/10.1145/2009916.2010018

[8] Hideki Kawai, Adam Jatowt, Katsumi Tanaka, Kazuo Kunieda, and Keiji Yamada. [n. d.]. ChronoSeeker: Search Engine for Future and Past Events. In *Proc. of the 4th ICUIMC* (New York, NY, USA, 2010) *(ICUIMC '10)*. Association for Computing Machinery, Article 25. https://doi.org/10.1145/2108616.2108647

[9] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach. arXiv:1907.11692 [cs.CL]

[10] Yoko Nakajima. [n. d.]. Automatic Extraction of References to Future Events from News Articles Using Semantic and Morphological Information. In *Proc. of the 24th IJCAI* (Buenos Aires, Argentina, 2015) *(IJCAI'15)*. AAAI Press, 4385–4386.

[11] Aiming Ni, Jinho D. Choi, Jason Shepard, and P. Wolff. [n. d.]. Computational Exploration to Linguistic Structures of Future: Classification and Categorization. In *North American Chapter of the Association for Computational Linguistics* (2015).

[12] Kira Radinsky, Sagie Davidovich, and Shaul Markovitch. [n. d.]. Predicting the News of Tomorrow Using Patterns in Web Search Queries. In *IEEE/WIC/ACM 2008* (2008), Vol. 1. 363–367. https://doi.org/10.1109/WIIAT.2008.215

[13] Kira Radinsky and Eric Horvitz. [n. d.]. Mining the Web to Predict Future Events. In *Proc. of the 6th WSDM* (New York, NY, USA, 2013) *(WSDM '13)*. Association for Computing Machinery, 255–264. https://doi.org/10.1145/2433396.2433431

[14] Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. arXiv:1908.10084 [cs.CL]

[15] Julius Steen and Katja Markert. 2019. Abstractive Timeline Summarization. In *Proc. of the 2nd Workshop on New Frontiers in Summarization*. Association for Computational Linguistics, Hong Kong, China, 21–31. https://doi.org/10.18653/v1/D19-5403

[16] Navya Yarrabelly and K. Karlapalem. [n. d.]. Extracting Predictive Statements with Their Scope from News Articles. In *ICWSM* (2018).

[17] Yi Yu, Adam Jatowt, Antoine Doucet, Kazunari Sugiyama, and Masatoshi Yoshikawa. [n. d.]. Multi-TimeLine Summarization (MTLS): Improving Timeline Summarization by Generating Multiple Summaries. In *Proc. of the 59th ACL and 11th IJCNLP* (Online, 2021-08). Association for Computational Linguistics, 377–387. https://doi.org/10.18653/v1/2021.acl-long.32